



TITLE:

Insular and hippocampal contributions to remembering people with an impression of bad personality.

AUTHOR(S):

Tsukiura, Takashi; Shigemune, Yayoi; Nouchi, Rui; Kambara, Toshimune; Kawashima, Ryuta

CITATION:

Tsukiura, Takashi ...[et al]. Insular and hippocampal contributions to remembering people with an impression of bad personality.. Social cognitive and affective neuroscience 2012, 8(5): 515-522

ISSUE DATE:

2012-02

URL:

<http://hdl.handle.net/2433/177006>

RIGHT:

© The Author (2012). Published by Oxford University Press.; この論文は出版社版ではありません。引用の際には出版社版をご確認ご利用ください。; This is not the published version. Please cite only the published version.

Insular and hippocampal contributions to remembering people with an impression of bad personality

**Takashi Tsukiura^{1,2}, Yayoi Shigemune^{1,2}, Rui Nouchi^{3,4},
Toshimune Kambara^{1,4}, Ryuta Kawashima^{1,3}**

¹Department of Functional Brain Imaging, Institute of Development, Aging and Cancer (IDAC), Tohoku University, Sendai 980-8575, Japan,

²Department of Cognitive and Behavioral Sciences, Graduate School of Human and Environmental Studies, Kyoto University, Kyoto 606-8501, Japan,

³Smart Ageing International Research Center, Institute of Development, Aging and Cancer (IDAC), Tohoku University, Sendai 980-8575, Japan, and

⁴Japan Society for the Promotion of Sciences, Tokyo 102-8472, Japan

Running title: fMRI study of memory for bad people

* Corresponding author:

Takashi Tsukiura, Ph.D.

Department of Cognitive and Behavioral Sciences, Graduate School of Human and Environmental Studies, Kyoto University

Yoshida-Nihonmatsu-Cho, Sakyo-ku, Kyoto 606-8501, Japan

Tel. +81-75-753-6792 Fax. +81-75-753-6716

E-mail: tsukiura.takashi.6c@kyoto-u.ac.jp

Our impressions of other people are formed mainly from the two possible factors of facial attractiveness and trustworthiness. Previous studies have shown the importance of orbitofrontal-hippocampal interactions in the better remembering of attractive faces, and psychological data have indicated that faces giving an impression of untrustworthiness are remembered more accurately than those giving an impression of trustworthiness. However, the neural mechanisms of the latter effect are largely unknown. To investigate this issue, we investigated neural activities with event-related fMRI while the female participants rated their impressions of the personalities of men in terms of trustworthiness. After the rating, memory for faces was tested to identify successful encoding activity. As expected, faces that gave bad impressions were remembered better than those that gave neutral or good impressions. In fMRI data, right insular activity reflected an increasing function of bad impressions, and bilateral hippocampal activities predicted subsequent memory success. Additionally, correlation between these insular and hippocampal regions was significant only in the encoding of faces associated with a bad impression. Better memory for faces associated with an impression of bad personality could reflect greater interaction between the avoidance-related insular region and the encoding-related hippocampal region.

Keywords: fMRI; face; hippocampus; insula; memory

INTRODUCTION

When we encounter someone for the first time, we rapidly form some impressions of that person. Impressions of people could be a very important cue in identifying and remembering other people who should be approached or avoided in social interaction. Although our impressions of other people are formed by face-based signals such as trustworthy, caring, responsible, emotionally stable, sociable, attractive, intelligent, confident, dominant, happy, aggressive, threatening, mean, or weird features (Oosterhof & Todorov, 2008; Todorov & Engell, 2008; Todorov et al., 2008b), two possible factors of the facial attractiveness and trustworthiness are important in modulating memory for faces (Tsukiura & Cabeza, 2011b). With regard to the facial attractiveness, there is functional neuroimaging evidence that interactions between the reward-related orbitofrontal cortex and the memory-related hippocampal region contribute to the better remembering of attractive faces compared to unattractive faces (Tsukiura & Cabeza, 2011a). With regard to the trustworthiness, psychological studies have reported that faces giving an impression of untrustworthiness are remembered more accurately than those giving a neutral or trustworthy impression (Mealey et al., 1996; Oda, 1997). An intriguing possibility is that memory enhancement for faces with an impression of untrustworthiness reflects an effect of brain regions associated with the processing of an impression of untrustworthiness, on brain regions associated with the processing of memory for faces. However, little is known about the neural mechanisms underlying this effect. The current functional MRI (fMRI) study investigated this hypothesis.

The important role of the insula in the processing of socially negative signals from faces or persons has been demonstrated by functional neuroimaging studies (Krendl et al., 2006; O'Doherty et al., 2003b; Phillips et al., 1997; Todorov et al., 2008a; Tsukiura & Cabeza, 2011b; Winston et al., 2002). For example, one fMRI study reported that insular activity increased during the processing of both unattractive faces and bad hypothetical actions of persons, compared to the processing of attractive faces and good hypothetical actions of persons (Tsukiura & Cabeza, 2011b). Insular activity for face-based social signals was also found during the processing of

untrustworthy personality from faces (Todorov et al., 2008a). These studies suggest that activity in the insular cortex could reflect the processing of negative signals from external features of faces, such as facial unattractiveness, as well as of internally generated negative feelings about other people, such as untrustworthiness. Moreover, activity in the insular cortex has been linked to the feeling of being hurt emotionally during a social interaction, or ‘social pain’ (Eisenberger et al., 2003; Sanfey et al., 2003). Thus, the insular cortex could contribute to the processing of the faces of people associated with bad impressions of untrustworthiness, which should be avoided in situations of social interaction, and could be modulated by negative feelings generated internally.

Another candidate associated with the processing of socially negative signals from faces is the amygdala, which has been demonstrated in cognitive neuroscience studies for brain-damaged patients and healthy participants. For example, patients with bilateral amygdala damage were impaired in discriminating untrustworthy- from trustworthy-looking faces (Adolphs et al., 1998). Functional neuroimaging studies have shown that the amygdala response to faces increased as their perceived untrustworthiness increased (Engell et al., 2007; Todorov et al., 2008a; Winston et al., 2002). Moreover, there is also functional neuroimaging evidence that the amygdala shows greater activity for both positive as well as negative social signals from faces in terms of facial attractiveness or trustworthiness (Said et al., 2009; Winston et al., 2007). Thus, activity in the amygdala during the processing of faces could be modulated by two possible factors of only negative signals or of both positive and negative signals from faces.

The enhancing effect of untrustworthy impressions on memory for faces could reflect an influence of the insular and amygdala regions on the medial temporal lobe (MTL) regions, and in particular on the hippocampus. There is functional neuroimaging evidence linking hippocampal activity to the processing of encoding and retrieval of memory details (Davachi, 2006; Diana et al., 2007). For example, activity in the hippocampus during encoding predicts subsequent retrieval with high confidence (Kim & Cabeza, 2007), which is a signature of vivid remembering or recollection. Hippocampal activity has also been demonstrated in the successful encoding of memory for

associations between multiple items (Achim & Lepage, 2005; Chua et al., 2007; Kirwan & Stark, 2004; Prince et al., 2005; Prince et al., 2007; Sperling et al., 2003; Summerfield et al., 2006), as well as between item and context (Davachi et al., 2003; Gold et al., 2006; Kensinger & Schacter, 2006; Ranganath et al., 2004; Sommer et al., 2005; Uncapher et al., 2006). Moreover, functional neuroimaging studies have reported that face memories, the processing of which involves the hippocampus, are enhanced by socially positive signals from faces, which involve the orbitofrontal cortex (Tsukiura & Cabeza, 2008, 2011a). Thus, better memory for faces with a subjective impression of untrustworthiness would be mediated by a modulatory effect of insular and/or amygdala activity on the hippocampal activity during the encoding of faces.

The design of this study is summarized in Figure 1. During the encoding phase, the neural activities of female participants were scanned with event-related fMRI during the processing of rating the goodness of the estimated personalities of men from faces with neutral facial expressions. The participants were instructed that the ratings should be based on a subjective impression of trustworthiness, but not be based on facial attractiveness, which was evaluated for each face after the experiment. In this phase, no instruction about a subsequent retrieval test was given to the participants, and therefore encoding was incidental. During the retrieval phase, participants were presented with old and new faces, for each of which they made a combined recognition/confidence judgment (definitely old, probably old, probably new, and definitely new). Brain activity in the encoding phase was analyzed by parametric modulation procedures, which employed two functions of bad personality impression, or untrustworthiness, and subsequent memory. On the basis of the aforementioned research, we made three predictions for encoding-related activations in this study. First, activity of the insular cortex and ventral amygdala would increase as a function of untrustworthy impression, and dorsal part of the amygdala would show greater activity for both trustworthy and untrustworthy impressions (Todorov, 2008). Second, activity of the hippocampus would predict subsequent recognition of faces with high confidence (recollection). Third, correlations between activities in impression-related insular/amygdala regions and memory-related

hippocampal regions would be significant for faces with an untrustworthy impression but not for faces with a neutral or trustworthy impression.

Fig. 1 about here

MATERIALS AND METHODS

Participants

Twenty-five right-handed, college-aged females were recruited from the Tohoku University community and were paid for their participation. All participants were healthy and native Japanese speakers, with no history of neurological or psychiatric disorders. The data from two participants were excluded from analyses because of equipment malfunction. Thus, our analyses included data from 23 participants with an average age of 21.8 years (s.d.=2.2). All participants gave informed consent to a protocol approved by the Institutional Review Board of Tohoku University School of Medicine.

Stimuli

The stimuli were 240 photos of Japanese male faces with neutral expressions. These faces were selected from a face database, which was used with the permission of the Softopia Japan Foundation (www.softopia.or.jp/rd/facedb.html). It is strictly prohibited to copy or reuse this database, or to distribute the facial data, without permission. Additionally, in order to have enough faces for the experiment, we also included photos of Japanese male fashion models found in online catalogs or magazines. All stimuli were converted into grayscale images with dimensions of 256 x 256 pixels on a white background. Although there would be a potential limitation, we employed female participants and male faces to compare findings in this study with those in our previous study of memory for attractive faces (Tsukiura & Cabeza, 2011a), which employed female participants and male faces. The 240 photos were divided into 180 photos that were presented during the encoding phase (old faces), and 60 photos that were used for new faces as distractors during the retrieval phase. Examples of stimuli are illustrated in Figure 1.

Experimental tasks

All participants performed both encoding and retrieval tasks, with a study-test delay of approximately 30 minutes. In these phases of encoding and retrieval, neural activities were measured by the event-related fMRI method. During both encoding and retrieval, each face was presented for 3500 ms and followed by a variable (500-6500 ms) fixation interval. During a block of the encoding phase (see Figure 1), participants were randomly presented with 60 male faces with neural expressions one by one, and were required to rate the personality goodness for each face based on subjective assessment using an 8-button response box (from 1=very bad to 8=very good). The participants were instructed that the personality goodness of the faces should be rated on the basis of trustworthiness, but not be evaluated on the basis of facial attractiveness. No reference was made to a subsequent memory test, and therefore encoding was incidental. This procedure was repeated three times in the encoding phase, so that participants completely learned 180 male faces. During a block of the retrieval phase after the encoding phase (see Figure 1), 60 old and 20 new faces were presented in random order. For each face, participants made a combined recognition/confidence judgment: 1=definitely old, 2=probably old, 3=probably new, and 4=definitely new. This procedure was also repeated three times in the retrieval phase, so that participants were tested on their memories for all 240 faces. Encoding and retrieval trials with no responses were excluded from fMRI analyses. After the retrieval phase, all participants viewed new faces presented in the retrieval phase, and were required to rate the personality goodness for each new face using the same instructions and procedures employed in the encoding phase. Additionally, to examine the effect of facial attractiveness on memory for faces, the attractiveness of all faces used in the experiment was evaluated by the same participants, using the 8-point scale (from 1=very unattractive to 8=very attractive). Behavioral responses in these tests outside the fMRI scanner were recorded using the keyboard of a Windows PC.

Encoding trials were divided according to the personality impression of trustworthiness and according to the subsequent memory performance. The personality impression was categorized into three conditions of Bad (levels 1-3), Neutral (levels 4-5), and Good (levels 6-8). This categorization was also

applied to the new faces presented only in the retrieval phase. The subsequent memory performance was also categorized into three conditions of subsequent misses (Miss), subsequent hits with low confidence (HL), and subsequent hits with high confidence (HH).

Image acquisition and data analysis

All MRI data acquisition was conducted with a 3-T Philips Achieva scanner. Stimuli were visually presented through a projector and back-projected onto a screen. Participants viewed the stimuli via a mirror attached with the head coil of an MRI scanner. Behavioral responses were recorded using an 8-button fiber-optic response box (Current Designs, Inc. Philadelphia, PA). Scanner noise was reduced with earplugs, and head motion was minimized using foam pads and a headband. Anatomical scans began by first acquiring a T1-weighted sagittal localizer series. Second, functional images were acquired utilizing echo-planar functional images (EPIs) sensitive to blood oxygenation level dependent (BOLD) contrast (64 x 64 matrix, TR=2000 ms, TE=30 ms, flip angle=70 degree, FOV=24 cm, 34 slices, 3.75 mm slice thickness). Finally, high-resolution T1-weighted structural images (MPRAGE, 240 x 240 matrix, TR=6.5 ms, TE=3 ms, FOV=24 cm, 162 slices, 1.0 mm slice thickness) were collected.

The preprocessing and statistical analyses for all images were performed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK) implemented in MATLAB (www.mathworks.com). In the preprocessing analysis, images were corrected for slice-timing and head motion, then spatially normalized into the Montreal Neurological Institute (MNI) template and spatially smoothed using a Gaussian kernel of 8 mm FWHM.

The fMRI analyses focused on data from the encoding phase. Retrieval-related activity will be reported elsewhere. Statistical fMRI analyses were performed first at the subject level and then at the group level. In subject-level fixed-effect analyses, trial-related activity was modeled by convolving a vector of trial onsets with a canonical hemodynamic response function (HRF) within the context of the General Linear Model (GLM). Confounding factors (head motion, magnetic field drift) were also included in the model. Activity associated with the processing of personality impressions and with

subsequent memory performance (Paller & Wagner, 2002) was identified using the parametric analyses with three levels. Badness-related activity in a subjective impression of trustworthiness was identified with a linear regressor (Bad=3, Neutral=2, Good=1), and encoding-related activity with a quasi-exponential regressor (Miss=1, HL=2, HH=9). The quasi-exponential regressor places a strong weight on encoding activity predicting subsequent recognition with high confidence, which is assumed to have a greater recollection component (Daselaar et al., 2006b; Tsukiura & Cabeza, 2011a). These procedures provided us with two contrast images reflecting badness-related and encoding-related activity for individual participants. Additionally, given evidence that the amygdala showed greater activity for both positive and negative social signals from faces in terms of facial attractiveness or trustworthiness (Said et al., 2009; Winston et al., 2007), we conducted exploratory analyses with a regressor (Bad=3, Neutral=1, Good=3) during encoding.

At the group-level random effect analysis, to identify badness-related and encoding-related activations that were consistent across participants, we conducted one-sample t-tests for two patterns of contrast images from the subject-level fixed-effect analyses. The statistical threshold at the voxel level was set at $P < 0.001$ and corrected for multiple comparisons at the cluster level (FWE) with minimum cluster size of 10 voxels ($P < 0.05$). Arousal-related activations for both bad and good impressions were also analyzed by a one-sample t-test for contrast images from the subject-level analysis at the same threshold. Additionally, to find regions reflecting both effects of bad personality impression and encoding, we conducted a conjunction analysis between these badness-related and encoding-related contrasts at the same threshold. If we identify regions showing significant activations in this conjunction analysis, activity in the regions would be modulated by both effects of a facial impression of bad personality and subsequent memory of faces. Within insular, MTL (hippocampus and parahippocampal gyrus), and amygdala regions-of-interest (ROIs), based on the a-priori hypothesis, we employed the small volume correction (SVC) method for these ROIs by using the same threshold. These ROIs were defined by the WFU PickAtlas (www.fmri.wfubmc.edu) and the AAL ROI package (Tzourio-Mazoyer et al.,

2002). All coordinates of activations were converted from MNI to Talairach space (Talairach & Tournoux, 1988) using MNI2TAL (www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml).

To investigate the effects of facial impression of bad personality on interaction between insular/amygdala and hippocampal regions that showed badness-related and encoding-related activity in the parametric modulation analyses, we conducted correlation analyses using the method employed in previous studies (Dolcos et al., 2004, 2005; Tsukiura & Cabeza, 2011a), where the authors investigated interacting activities between memory-related MTL and emotion-related regions such as the amygdala or orbitofrontal cortex. From the activation cluster of the insular cortex, amygdala, and hippocampus showing badness-related and encoding-related activity, we extracted activation levels (effect sizes) for HH trials in each of three impression conditions of personality (Bad, Neutral, and Good). Using these data of insular, amygdala, and hippocampal activities, we computed a separate Pearson correlation for each impression condition.

RESULTS

Behavioral data

Table 1 shows proportions of accuracy (%) and RTs (ms) during retrieval, and RTs (ms) of goodness ratings during encoding. The impression of untrustworthiness affected recognition accuracy during retrieval, but the enhancement was identified only in recognition with high confidence. As illustrated by Figure 2, an analysis of variance (ANOVA) for recognition accuracy (proportions of hits vs. false alarms) on high-confidence responses showed a significant effect of an impression factor ($F=13.60$, $P<0.01$), and post-hoc tests showed that the recognition accuracy for faces with an impression of bad personality was significantly higher than that for faces with an impression of neutral personality ($P<0.01$) and of good personality ($P<0.01$). In contrast, the effect of an impression factor on proportions of total recognition accuracy (hits vs. total false alarms on both confidence levels) was not significant ($F=2.19$, $P=0.12$). The findings suggest that a subjective

impression of untrustworthiness could enhance the remembering who should be avoided on a later occasion, rather than the simple recognition of old faces.

To examine further the finding that the effect of an impression factor on recognition performance is not related to the effect of a facial attractiveness factor, we conducted another ANOVA for proportions of recognition accuracy (proportions of hits vs. false alarms). According to the subjective rating scores of facial attractiveness for facial stimuli, facial attractiveness was categorized into three conditions of Unattractive (levels 1-3), Neutral (levels 4-5), and Attractive (levels 6-8). An ANOVA for recognition accuracy (proportions of hits vs. false alarms) on high-confidence responses (Unattractive: 0.35, Neutral: 0.30, Attractive: 0.33) showed no significant effect of a facial attractiveness factor ($F=0.40$, $P=0.68$). In proportions of total recognition accuracy (hits vs. false alarms on both confidence levels), we also found no significant effect of facial attractiveness ($F=0.30$, $P=0.75$). Additionally, we analyzed correlations between the goodness and attractiveness judgments, using mean rating scores of facial goodness and attractiveness in each face. This analysis showed a significantly positive correlation ($r=0.63$, $P<0.01$). These findings are consistent with our assumption that the memory enhancement by a subjective impression of untrustworthiness is independent of the effects of facial attractiveness, while the rating scores of facial goodness and attractiveness were correlated each other.

Fig. 2 about here

Regarding RTs (ms) of impression ratings during encoding (see Table 1), a two-way ANOVA with factors of subsequent recognition performance (HH, HL, and Miss) and impression (Bad, Neutral, and Good) showed a significant main effect of impression ($F=10.03$, $P<0.01$), where rating faces with a bad or good impression was significantly faster than rating faces with a neutral impression (both $P<0.01$). However, we did not find a significant main effect of subsequent recognition performance ($F=0.29$, $P=0.75$) and interaction between factors of subsequent recognition performance and impression ($F=1.18$, $P=0.33$). Faster RTs for faces with a untrustworthy or trustworthy impression than for faces with a neutral impression suggest that rating impressions of estimated personality was more difficult in the middle range of the estimated personality scale. It is worth pointing out that better memory

for faces with an untrustworthy impression cannot be attributed to longer encoding times.

RTs (ms) for successful recognition with high confidence tended to be faster than for successful recognition with low confidence or for missed recognition (see Table 1). A two-way ANOVA with factors of recognition performance (HH, HL and Miss) and impression (Bad, Neutral and Good) on the RTs showed a significant main effect of recognition performance ($F=36.59$, $P<0.01$), with HH being significantly faster than HL ($P<0.01$) and Miss ($P<0.01$), and with HL being significantly faster than Miss ($P<0.05$). However, a main effect of impression ($F=2.33$, $P=0.11$) and interaction between recognition performance and impression factors ($F=1.67$, $P=0.16$) were not significant.

[Table 1 about here](#)

fMRI data

Confirming our first prediction, insular activity increased as a function of impression of untrustworthiness (see Figure 3). As indicated in Table 2, significantly linear increases of activity as a function of an untrustworthy impression were also found in other regions, including the precentral gyrus, putamen, and cerebellar hemisphere.

[Table 2 about here](#)

Confirming our second prediction, hippocampal activity during encoding predicted subsequent recollection of faces. This activation was found in the bilateral anterior hippocampi including the border with the amygdala. In keeping with the quasi-exponential regressor, the activation showed a sharp increase for HH (see Figure 4), which is the signature of recollection-related activity (Daselaar et al., 2006b; Tsukiura & Cabeza, 2011a). Activation profiles in the bilateral hippocampi are summarized in Table 2.

To find regions reflecting both effects of bad impression and encoding, we conducted a conjunction analysis between the two parametric regressors. This analysis showed no significant activation in any brain region (see Table 2). In addition to the badness-related and encoding-related analyses, we conducted a parametric modulation analysis to identify brain regions associated with impressions of both bad and good personalities. Arousal-related activations

reflecting these impressions were not identified in any region including amygdala ROI.

Finally, confirming our third prediction, during the encoding of faces recognized with high confidence, interaction between activities in the right insula and left hippocampus identified in the parametric analyses was significant for faces with an untrustworthy impression but not for faces with a neutral or trustworthy impression (see Figure 5). We computed correlations between the right insula and left hippocampus (Pearson) in each condition of impression during HH, and the correlation coefficient for faces with a bad impression was significant ($r=0.56$, $P<0.01$), whereas those for faces with a neutral impression ($r=0.06$, $P=0.80$) or a good impression ($r=0.34$, $P=0.12$) were not. To confirm that the significant correlation for faces with a bad impression was not driven by one potential outlier (see Figure 5), we redid the correlation analysis without this point, and found that the correlation remained significant ($r=0.42$, $P=0.05$). Correlations between the right insula and right hippocampus were not significant in all conditions of facial impression (Bad: $r=0.19$, $P=0.38$; Neutral: $r=0.06$, $P=0.78$; Good: $r=0.25$, $P=0.26$).

DISCUSSION

Three main findings emerged from this study. First, insular activity during encoding increased as a function of impression of untrustworthiness. Second, hippocampal activity during encoding predicted subsequent memory accuracy and confidence. Finally, interaction between activities in the insula and hippocampus during the encoding of faces was modulated by an impression of estimated personality, and was significant only for faces with an untrustworthy (bad) impression. Taken together, these findings indicate that memory for faces with a subjective impression of bad personality could be enhanced by the modulatory effect of insular activity on hippocampal activity during encoding. Each of these three findings is discussed in separate sections below.

Activity as a function of an impression of bad personality

The first main finding of our study was that activity in the insular cortex increased as a function of untrustworthy (bad) impression (see Figure 3). This finding is consistent with functional neuroimaging evidence that this region showed greater activity during the processing of bad impressions for persons (Todorov et al., 2008a; Tsukiura & Cabeza, 2011b). The insular activity has also been identified during the processing of unattractive faces (Krendl et al., 2006; O'Doherty et al., 2003b; Tsukiura & Cabeza, 2011b) and facial expression of disgust (Phillips et al., 1997). The insular activity could reflect the processing of socially negative signals from both facial unattractiveness and untrustworthiness.

Fig. 3 about here

The insular responses to socially negative signals from faces or persons could be explained in terms of the role of this region in the processing of negative social situations, including social exclusion (Eisenberger et al., 2003), unfairness (Sanfey et al., 2003), and unreciprocated cooperation (Rilling et al., 2008). Functional neuroimaging studies have linked insular activations to the processing of punishment (O'Doherty et al., 2003a), emotions of disgust and fear (Phan et al., 2002), pain (Critchley et al., 2000), and aversive conditioning (Seymour et al., 2004). Moreover, there is functional neuroimaging evidence that insular activity was identified during the viewing of untrustworthy faces, when the processing of untrustworthiness of externally neutral faces was induced by internal feelings of the participants (Todorov et al., 2008a; Winston et al., 2002). The insular activity in our study suggests that activity in this region during the processing of bad impressions for others could raise the aversive feelings or emotional hurt (social pain) in our mind, and that the negative feeling could produce the avoiding behaviors for persons in the situation of social interaction.

Insular activity in our study increased as a function of untrustworthy impressions from faces, but the activity seemed to be negative compared to the baseline activity. This pattern of insular activity was also identified in our previous study, where the insular cortex showed increasing activity during the processing of both unattractive faces and sentences describing bad hypothetical actions of persons (Tsukiura & Cabeza, 2011b). This insular

activity may be interpreted in the context of social interaction. For example, one fMRI study reported a greater activity in the insula combined with an increased deactivation in the default mode network (Raichle et al., 2001) when participants were imitated, reflecting the social engagement with others required by social interaction (Guionnet et al., in press). However, the present finding is not enough to support this idea, and then further studies would be required.

Contrary to our expectation, the amygdala showed no increasing activity specific to face-based negative signals, or to both face-based negative and positive signals. This finding is inconsistent with cognitive neuroscience studies linking this region to the processing of untrustworthy faces (Adolphs et al., 1998; Engell et al., 2007; Todorov et al., 2008b; Winston et al., 2002), or of both untrustworthy and trustworthy faces (Said et al., 2009). The fact that amygdala activity in our study was not modulated by the processing of face trustworthiness suggests that the individual judgments for facial impressions may not be a good predictor to identify amygdala activations. For example, one fMRI study reported that judgments averaged across participants are a better predictor of amygdala activity than individual judgments during the processing of face trustworthiness (Engell et al., 2007). In the present study, we employed rating values of individual judgments as a regressor to find brain activation patterns, which might disable us to find amygdala activations.

Activity as a function of subsequent memory

The second main finding of our study was that activity in the hippocampus increased as a function of subsequent memory accuracy and confidence (see Figure 4). This finding suggests that hippocampal activity during encoding could predict subsequent retrieval success with high confidence, which reflects the vivid remembering process known as recollection.

Fig. 4 about here

The involvement of the hippocampus in predicting subsequent recollection is consistent with functional neuroimaging evidence that activity of this region reflects a function of subsequent retrieval success with high confidence (Daselaar et al., 2006b; Kim & Cabeza, 2007; Tsukiura & Cabeza,

2011a). For example, one fMRI study found that by applying a quasi-exponential function, hippocampal activity during the encoding of faces predicted subsequent retrieval success with high confidence (Tsukiura & Cabeza, 2011a). In addition, functional neuroimaging studies have consistently reported that the hippocampus shows significant activations during the successful encoding of relational memories (Achim & Lepage, 2005; Chua et al., 2007; Kirwan & Stark, 2004; Prince et al., 2005; Prince et al., 2007; Sperling et al., 2003; Summerfield et al., 2006; Tsukiura & Cabeza, 2008) or memories for contextual details (Davachi et al., 2003; Gold et al., 2006; Kensinger & Schacter, 2006; Ranganath et al., 2004; Sommer et al., 2005; Uncapher et al., 2006). The encoding-related activity in the hippocampus could be modulated by subsequent recollection processes, even when the encoding procedure is incidental.

Moreover, the present data revealed that the hippocampus predicted subsequent memory success with high confidence, whereas the parahippocampal gyrus, which was included in ROIs defined in the statistical analysis of our study, showed no significant activation as a function of subsequent memory. The absence of parahippocampal activity could be explained by the model of dissociable roles within the MTL regions between recollection and familiarity. For example, one fMRI study found triple dissociation within the MTL regions during the processing of recollection, familiarity, and novelty (Daselaar et al., 2006a). In the same study, the recollection process was associated with the posterior half of the hippocampus, the familiarity process with the posterior parahippocampal gyrus, and the novelty process with the anterior half of the hippocampus and the rhinal regions. Similar dissociable patterns within the MTL memory system have been shown in recent review studies (Davachi, 2006; Diana et al., 2007). The present MTL activation patterns, where we identified significant activations in the hippocampus but not in the parahippocampal gyrus, would reflect possible dissociable roles within the MTL regions during encoding.

Insular-hippocampal interaction in remembering faces associated with a subjective impression of bad personality

The third main finding of our study was that interaction between insular and hippocampal activities was significant only in the successful encoding of faces with an untrustworthy (bad) impression, but not in the successful encoding of faces with a neutral or trustworthy (good) impression (see Figure 5). This finding suggests that a modulatory effect of the insular cortex, which is involved in the processing of face-based negative signals, on the hippocampus, which is involved in the subsequent recollection process during encoding, could explain the enhancing effect of an untrustworthy (bad) impression on memory for faces.

Fig. 5 about here

The finding of significant interaction between the insula and hippocampus during the encoding of untrustworthy faces is consistent with functional neuroimaging evidence that activity in the insula and hippocampus was greater during the encoding of emotional stimuli than of neutral stimuli (Rasch et al., 2009; St Jacques et al., 2009). For example, the subsequent memory effect (“difference due to memory” or Dm) was larger for emotional than for neutral pictures, where greater activities in the insular and hippocampal regions were associated with enhanced Dm for emotional pictures (Rasch et al., 2009). Another fMRI study reported that, during the encoding of negative pictures, greater insular activity was observed in young participants, who exhibited a larger effect of emotional memory enhancement than older participants (St Jacques et al., 2009). The present finding that faces with an untrustworthy impression were remembered better than faces with a neutral or trustworthy impression, and the fact that the correlation between insular and hippocampal activities was significant only in the successful encoding of faces with an untrustworthy impression, suggest that insular responses to negative stimuli could modulate encoding-related hippocampal activity. However, further analyses would be required to clarify the importance of this interaction during the encoding of untrustworthy faces, because the statistical power may not be enough in our study.

In addition to stimulus-dependent emotional memory, significant activations in the insula and hippocampus were identified in the processing of

internally generated disgust feelings via autobiographical memory (Fitzgerald et al., 2004). Previous neuroimaging data have shown that insular activity is identified in the processing of the feeling of being hurt emotionally during a social interaction (Eisenberger et al., 2003; Sanfey et al., 2003), and in anticipating outcomes of subsequent aversive stimuli in the context of uncertainty (Sarinopoulos et al., 2010). In our study, given that all facial stimuli employed had neutral expressions, and that the effect of facial attractiveness on memory for faces was not significant, impressions from faces that had no external value of emotion would be induced internally in participants. These findings suggest that not only emotional items presented externally but also internally generated negative feelings could be processed in the insular cortex, which could affect the hippocampal memory system. Given that the recollection of faces was enhanced by an impression of untrustworthiness, whereas the enhancement was not found in the simple recognition of faces, insular-hippocampal interaction, which was enhanced during the encoding of faces with an impression of untrustworthiness, could contribute to remembering people who should be avoided in the uncertain situation of social interaction.

CONCLUSION

Using event-related fMRI, we investigated the effect of face-based impressions of trustworthiness on brain activity during the successful encoding of faces. Insular activity increased as a function of an untrustworthy impression, where the activity was greater during the processing of faces with an untrustworthy (bad) impression than of faces with a neutral or trustworthy (good) impression. Activity in the hippocampus increased as a function of subsequent memory success, where the activity predicted subsequent successful recollection of faces. In addition, interaction between these insular and hippocampal activities during the encoding of faces recollected successfully in the retrieval test was significant only for faces with an untrustworthy impression, but not for faces with a neutral or trustworthy impression. Taken together with our behavioral data, in which recollection memory performance for faces with an untrustworthy impression was better than that for faces with a neutral or trustworthy impression, these findings suggest that insular-

hippocampal interaction could contribute to remembering people who should be avoided in social interactions.

Funding

This research was funded by the Cabinet Office, Government of Japan, through its “Funding Program for Next Generation World-Leading Researchers”. Part of this study was also supported by Grant-in-Aid for Scientific Research on Innovative Areas, “Face perception and Recognition” (21119503) through the Ministry of Education, Science, Sports and Culture, Japan.

REFERENCES

- Achim, A. M., & Lepage, M. (2005). Neural correlates of memory for items and for associations: an event-related functional magnetic resonance imaging study. *J Cogn Neurosci*, *17*, 652-667.
- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470-474.
- Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2007). Evidence for a specific role of the anterior hippocampal region in successful associative encoding. *Hippocampus*, *17*, 1071-1080.
- Critchley, H. D., Elliott, R., Mathias, C. J., & Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *J Neurosci*, *20*, 3033-3040.
- Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2006a). Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. *J Neurophysiol*, *96*, 1902-1911.
- Daselaar, S. M., Fleck, M. S., Dobbins, I. G., Madden, D. J., & Cabeza, R. (2006b). Effects of healthy aging on hippocampal and rhinal memory functions: an event-related fMRI study. *Cereb Cortex*, *16*, 1771-1782.
- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Curr Opin Neurobiol*, *16*, 693-700.
- Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc Natl Acad Sci U S A*, *100*, 2157-2162.
- Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends Cogn Sci*, *11*, 379-386.
- Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron*, *42*, 855-863.
- Dolcos, F., LaBar, K. S., & Cabeza, R. (2005). Remembering one year later: role of the amygdala and the medial temporal lobe memory system in

- retrieving emotional memories. *Proc Natl Acad Sci U S A*, 102, 2626-2631.
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302, 290-292.
- Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *J Cogn Neurosci*, 19, 1508-1519.
- Fitzgerald, D. A., Posse, S., Moore, G. J., Tancer, M. E., Nathan, P. J., & Phan, K. L. (2004). Neural correlates of internally-generated disgust via autobiographical recall: a functional magnetic resonance imaging investigation. *Neurosci Lett*, 370, 91-96.
- Gold, J. J., Smith, C. N., Bayley, P. J., et al. (2006). Item memory, source memory, and the medial temporal lobe: concordant findings from fMRI and memory-impaired patients. *Proc Natl Acad Sci U S A*, 103, 9351-9356.
- Guionnet, S., Nadel, J., Bertasi, E., Sperduti, M., Delaveau, P., & Fossati, P. (in press). Reciprocal Imitation: Toward a Neural Basis of Social Interaction. *Cereb Cortex*.
- Kensinger, E. A., & Schacter, D. L. (2006). Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. *J Neurosci*, 26, 2564-2570.
- Kim, H., & Cabeza, R. (2007). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cereb Cortex*, 17, 2143-2150.
- Kirwan, C. B., & Stark, C. E. L. (2004). Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. *Hippocampus*, 14, 919-930.
- Krendl, A. C., Macrae, C. N., Kelley, W. M., Fugelsang, J. A., & Heatherton, T. F. (2006). The good, the bad, and the ugly: an fMRI investigation of the functional anatomic correlates of stigma. *Soc Neurosci*, 1, 5-15.
- Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiol*, 17, 119-128.

- O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003a). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J Neurosci*, *23*, 7931-7939.
- O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003b). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, *41*, 147-155.
- Oda, R. (1997). Biased face recognition in the Prisoner's Dilemma Game. *Evol Hum Behav*, *18*, 309-315.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proc Natl Acad Sci U S A*, *105*, 11087-11092.
- Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends Cogn Sci*, *6*, 93-102.
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, *16*, 331-348.
- Phillips, M. L., Young, A. W., Senior, C., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*, 495-498.
- Prince, S. E., Daselaar, S. M., & Cabeza, R. (2005). Neural correlates of relational memory: successful encoding and retrieval of semantic and perceptual associations. *J Neurosci*, *25*, 1203-1210.
- Prince, S. E., Tsukiura, T., & Cabeza, R. (2007). Distinguishing the neural correlates of episodic memory encoding and semantic memory retrieval. *Psychol Sci*, *18*, 144-151.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc Natl Acad Sci U S A*, *98*, 676-682.
- Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*, 2-13.
- Rasch, B., Spalek, K., Buholzer, S., et al. (2009). A genetic variation of the noradrenergic system is related to differential amygdala activation

- during encoding of emotional memories. *Proc Natl Acad Sci U S A*, *106*, 19191-19196.
- Rilling, J. K., Goldsmith, D. R., Glenn, A. L., et al. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia*, *46*, 1256-1266.
- Said, C. P., Baron, S. G., & Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information. *J Cogn Neurosci*, *21*, 519-528.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*, 1755-1758.
- Sarinopoulos, I., Grupe, D. W., Mackiewicz, K. L., et al. (2010). Uncertainty during anticipation modulates neural responses to aversion in human insula and amygdala. *Cereb Cortex*, *20*, 929-940.
- Seymour, B., O'Doherty, J. P., Dayan, P., et al. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*, 664-667.
- Sommer, T., Rose, M., Glascher, J., Wolbers, T., & Buchel, C. (2005). Dissociable contributions within the medial temporal lobe to encoding of object-location associations. *Learn Mem*, *12*, 343-351.
- Sperling, R., Chua, E., Cocchiarella, A., et al. (2003). Putting names to faces: successful encoding of associative memories activates the anterior hippocampal formation. *Neuroimage*, *20*, 1400-1410.
- St Jacques, P. L., Dolcos, F., & Cabeza, R. (2009). Effects of aging on functional connectivity of the amygdala for subsequent memory of negative pictures: a network analysis of functional magnetic resonance imaging data. *Psychol Sci*, *20*, 74-84.
- Summerfield, C., Greene, M., Wager, T., Egner, T., Hirsch, J., & Mangels, J. (2006). Neocortical connectivity during episodic memory formation. *PLoS Biol*, *4*, e128.
- Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotactic Atlas of the Human Brain: 3-dimensional Proportional System: An Approach to Cerebral Imaging*. Stuttgart: Georg Thieme.

- Todorov, A. (2008). Evaluating faces on trustworthiness: an extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Ann N Y Acad Sci*, 1124, 208-224.
- Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008a). Evaluating face trustworthiness: a model based approach. *Soc Cogn Affect Neurosci*, 3, 119-127.
- Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Soc Cogn Affect Neurosci*, 3, 303-312.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008b). Understanding evaluation of faces on social dimensions. *Trends Cogn Sci*, 12, 455-460.
- Tsukiura, T., & Cabeza, R. (2008). Orbitofrontal and hippocampal contributions to memory for face-name associations: the rewarding power of a smile. *Neuropsychologia*, 46, 2310-2319.
- Tsukiura, T., & Cabeza, R. (2011a). Remembering beauty: roles of orbitofrontal and hippocampal regions in successful memory encoding of attractive faces. *Neuroimage*, 54, 653-660.
- Tsukiura, T., & Cabeza, R. (2011b). Shared brain activity for aesthetic and moral judgments: implications for the Beauty-is-Good stereotype. *Soc Cogn Affect Neurosci*, 6, 138-148.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15, 273-289.
- Uncapher, M. R., Otten, L. J., & Rugg, M. D. (2006). Episodic encoding is more than the sum of its parts: an fMRI investigation of multifeatureal contextual encoding. *Neuron*, 52, 547-556.
- Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I., & Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, 45, 195-206.
- Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat Neurosci*, 5, 277-283.

Table 1 Behavioral results

	Bad (s.d.)	Neutral (s.d.)	Good (s.d.)
<i>Accuracy (%) of recognition responses</i>			
HH	38.3 (12.8)	29.9 (12.7)	28.8 (11.8)
FH	4.4 (5.2)	5.3 (6.9)	6.1 (9.3)
HH - FH	33.9 (11.6)	24.5 (9.8)	22.7 (10.9)
Total H	74.7 (12.0)	72.1 (10.4)	71.4 (11.0)
Total F	31.9 (18.0)	32.1 (22.5)	35.8 (18.0)
Total H - Total F	42.8 (17.0)	40.0 (17.8)	35.6 (15.1)
<i>RT (ms) during encoding</i>			
HH	1772.1 (269.3)	1912.2 (307.0)	1819.7 (341.5)
HL	1773.4 (293.6)	1906.9 (319.7)	1778.6 (310.1)
Miss	1807.9 (302.6)	1879.8 (285.5)	1800.3 (284.0)
<i>RT (ms) during retrieval</i>			
HH	1446.6 (287.1)	1480.1 (269.7)	1511.9 (293.6)
HL	1815.4 (287.7)	1812.0 (291.1)	1794.6 (312.3)
Miss	1899.7 (417.4)	1980.8 (390.7)	1987.6 (480.3)

H: hits, F: false alarms, HH: hits with high confidence, HL: hits with low confidence, FH: false alarms with high confidence, FL: false alarms with low confidence, Miss: misses, s.d.: standard deviation

Table 2 Regions showing parametric effects of facial impression of badness and subsequent memory

Regions	L/R	BA	Coordinates			Z value
			x	y	z	
<i>Linear increases with bad impression of faces</i>						
Insula	R	13	42	-13	13	5.50
Precentral gyrus	R	4	34	-18	58	7.43
Putamen	R		30	-14	0	6.30
Cerebellar hemisphere	L		-18	-52	-21	7.41
<i>Quasi-exponential increases with subsequent memory</i>						
Hippocampus	R		23	-4	-23	4.69
Hippocampus	L		-18	-12	-26	4.70
<i>Conjunction between both effects of bad impression and subsequent memory</i>						
No significant activation						

Note: R, right; L, left; BA, Brodmann area.

Figure legends

Fig. 1 Task paradigm. During encoding, female participants were required to rate the facial impression of badness for male faces using an eight-point scale (from 1: very bad, to 8: very good). During retrieval, previously studied and new faces were presented one by one. For each face, participants indicated whether the face was judged as (1) a studied face with high confidence (definitely old: DO), (2) a studied face with low confidence (probably old: PO), (3) an unstudied face with low confidence (probably new: PN), or (4) an unstudied face with high confidence (definitely new: DN).

Fig. 2 Proportion of hits vs. false alarms with high confidence. HH: high-confidence hits. FH: high-confidence false alarms. Error bars represent standard error. $**P < 0.01$.

Fig. 3 Badness-related activity and activation profile in the right insular cortex. Error bars represent standard error.

Fig. 4 Encoding-related activity and activation profile in the bilateral hippocampi. Error bars represent standard error. HH: high-confidence hits, HL: low-confidence hits, Miss: misses.

Fig. 5 Correlation between badness-related insular activity and encoding-related hippocampal (left) activity, separately for faces with a bad impression (green), faces with a neutral impression (blue), and faces with a good impression (orange). Insular and hippocampal activations were significantly correlated for faces with a bad impression ($**P < 0.01$), but not for faces with a neutral or good impression.

Fig. 1

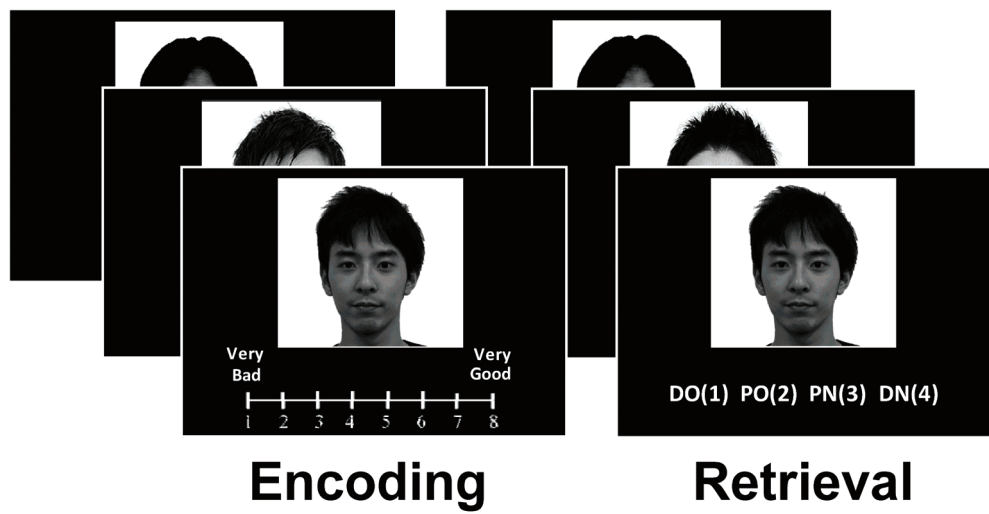


Fig. 2

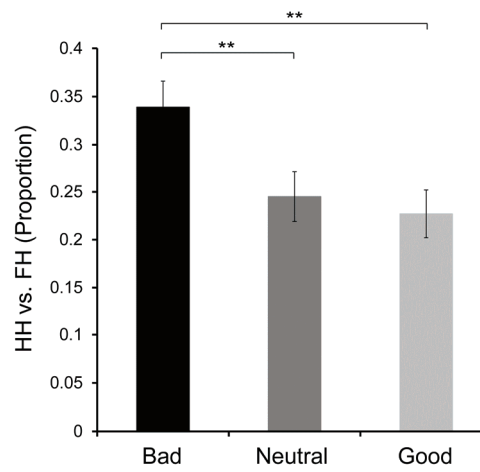


Fig. 3

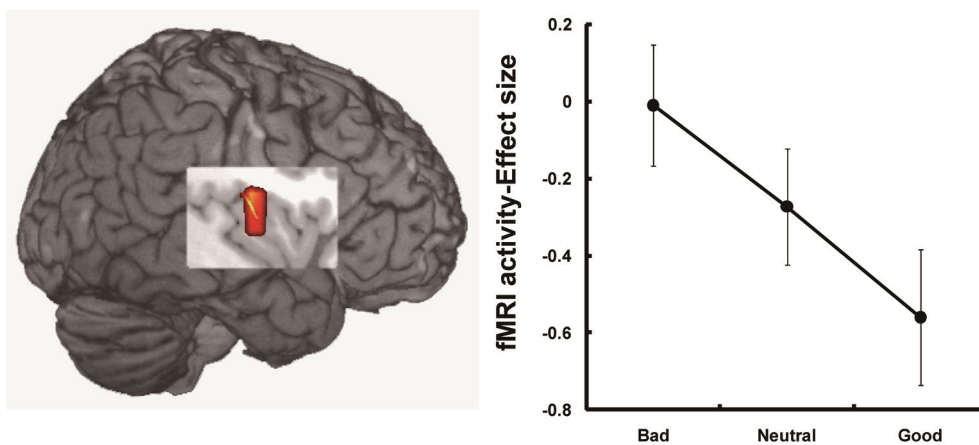


Fig. 4

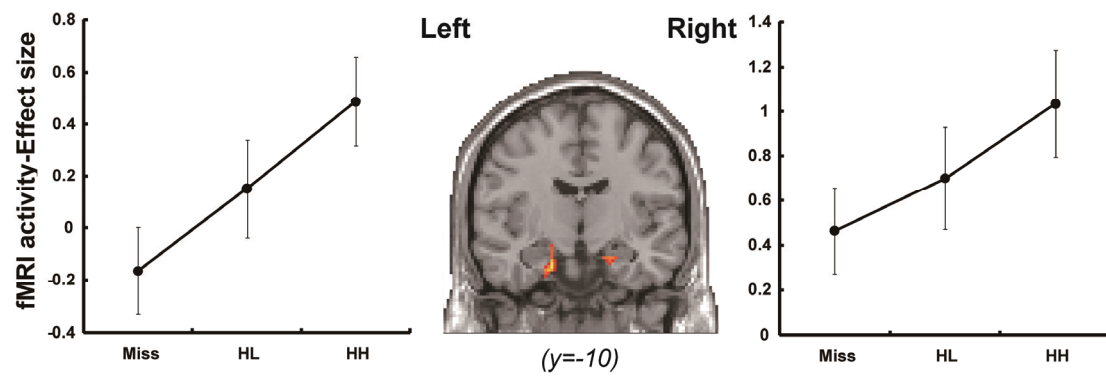


Fig. 5

